

NETWORK ROUTING USING INDIRECT NEXT HOP DATA

TECHNICAL FIELD

[0001] The invention relates to computer networks and, more particularly, to techniques for routing packets within computer networks.

BACKGROUND

[0002] A computer network is a collection of interconnected computing devices that can exchange data and share resources. In a packet-based network, such as the Internet, the computing devices communicate data by dividing the data into small blocks called packets, which are individually routed across the network from a source device to a destination device. The destination device extracts the data from the packets and assembles the data into its original form. Dividing the data into packets enables the source device to resend only those individual packets that may be lost during transmission.

[0003] Certain devices, referred to as routers, maintain routing information that describes routes through the network. A "route" can generally be defined as a path between two locations on the network. Upon receiving an incoming packet, the router examines information within the packet to identify the destination for the packet. Based on the destination, the router forwards the packet in accordance with the routing information.

[0004] Conventional routers often maintain the routing information in the form of one or more routing tables. The form and contents of the routing tables often depends on the routing algorithm implemented by the router. Common routing algorithms include distance vector routing algorithms and path vector routing algorithms. Many of these algorithms make use of the concept of a "hop," which refers to a connection between two devices. Consequently, the distance between two devices is often measured in hops. Furthermore, in reference to routing a packet, the "next hop" from a network router typically refers to a neighboring device along a given route.

[0005] The physical connection between two devices on a network is generally referred to as a link. Many conventional computer networks, including the Internet, are designed to dynamically reroute data packets in the event of a topology change, such as a link failure. Upon a topology change, the routers transmit new connectivity information to neighboring

devices, allowing each device to update its local routing information. Links can fail for any number of reasons, such as failure of the physical infrastructure between the devices, or failure of the devices interfacing with the link. The size and complexity of routing information maintained by routers within large networks can be significant. As a result, updating the routing information due to changes in network topology can consume considerable computing resources and substantially delay rerouting of packets.

SUMMARY

[0006] In general, the invention provides for increased efficiency in updating routing information after a change in network topology, such as link failure. According to the principles of the invention, a router maintains routing information that makes use of indirect references to identify the appropriate next hop for each route. In other words, intermediate data structures are introduced, referred to herein as indirect next hop data, between the routing information and the next hop information. The routing information is structured such that routes having the same next hop use indirect next hop data structures to point to common next hop data.

[0007] The invention offers many advantages, including reducing the impact and latency of network topology changes by reducing the computer resources required to update the routing information. In particular, in response to a change in network topology, the router need not change all of the affected routes, only the common next hop data referenced by the intermediate data structures. The router can, for example, overwrite the common next hop data with new next hop data. In this fashion, the router can effectively update a large number of routes, and thereby dynamically reroute packets, with minimal changes to the routing information.

[0008] In one embodiment, the invention is directed to a method including routing packets within a network using indirect next hop data that associates a plurality of routes with a common portion of next hop data.

[0009] In another embodiment, the invention is directed to a method including storing route data representing routes within a computer network, and storing next hop data representing network devices that neighbor a network router. The method further includes storing indirect next hop data that maps at least a subset of the routes represented by the route

data to a common portion of the next hop data. The route data may be stored as a radix tree, and the indirect next hop data may be stored as data pointers within leaf nodes of the radix tree.

[0010] In another embodiment, the invention is directed to a router comprising a routing engine to store routing information representing a topology of a network. The router further comprises a packet forwarding engine to store packet forwarding information in accordance with the routing information, the packet forwarding information including (i) route data representing destinations within a computer network, (ii) next hop data representing interfaces to neighboring network devices, and (iii) indirect next hop data that maps a subset of the routes represented by the route data to a common portion of the next hop data.

[0011] In another embodiment, the invention is directed to a router comprising a computer-readable medium to store: (i) route data representing routes within a computer network, (ii) next hop data representing neighboring network devices, and (iii) indirect next hop data that maps at least a subset of route data to a common portion of the next hop data.

[0012] In another embodiment, the invention is directed to a computer-readable medium containing data structures. The data structures include a first data structure to store route data representing destinations within a computer network, a second data structure to store next hop data representing interfaces to neighboring network devices, and a set of indirect next hop data that map at least a subset of the route data to a common portion of the next hop data.

[0013] The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0014] FIG. 1 is a block diagram illustrating an example router configured consistent with the principles of the invention.

[0015] FIG. 2A is a block diagram illustrating data structures for one exemplary arrangement of routing information making use of indirect next hop data.

[0016] FIG. 2B is a block diagram illustrating data structures for another exemplary arrangement of routing information making use of indirect next hop data.

[0017] FIG. 3 is a block diagram illustrating an example data structure for resolving next hop data to interface ports or other next hop actions.

[0018] FIG. 4 is a block diagram illustrating another example router configured consistent with the principles of the invention.

[0019] FIG. 5 is a flow chart illustrating an example operation of a router making use of indirect next hop data.

[0020] DETAILED DESCRIPTION

[0021] FIG. 1 is a block diagram illustrating an example router 4 configured consistent with the principles of the invention. In the exemplary embodiment illustrated in FIG. 1, router 4 includes one or more interface cards (IFCs) 6 for sending and receiving packets using network links 12 and 13. IFCs 6 are typically coupled to network links 12, 13 via a number of interface ports. In general, router 4 receives inbound packets from network links 12, determines destinations for the received packets, and outputs the packets on network links 13 based on the destinations.

[0022] Router 4 includes a routing engine 8 that maintains routing information 10, which describes a topology of a network and, in particular, the routes through the network. Routing information 10 may include, for example, route data 14 that describes various routes within the network, and corresponding next hop data 18 indicating appropriate neighboring devices within the network for each of the routes.

[0023] Routing information 10 may associate each next hop with one of network links 13 or IFCs 6. In particular, upon receiving an inbound packet, routing engine 8 determines a route within route data 14 for the inbound packet, and examines next hop data 18 of routing information 10 to identify a next hop for the packet. Based on the identified next hop, routing engine 8 determines an interface port associated with the next hop, and forwards the inbound packet to the appropriate IFC 6 for transmission. The architecture of router 4 illustrated in FIG. 1 is for exemplary purposes only. The invention is not limited to this architecture. In other embodiments, router 4 may be configured in a variety of ways. In one embodiment, for example, routing engine 8 and its corresponding functionality may be replicated and incorporated directly within IFCs 6.

[0024] According to the principles of the invention, routing information 10 may make use of indirect references to associate routes with corresponding next hops. In other words, routing information 10 may use intermediate data structures, referred to herein as indirect next hop data 16, that maps route data 14 to next hop data 18. In particular, indirect next hop data 16 is structured such that routes that make use of the same next hop from router 4 reference a common portion of next hop data 18. In this manner, router 4 need not maintain separate next hop data for each individual route. In addition, routing information 10 may maintain references that bypass indirect next hop data 16, and associate route data 14 directly with next hop data 18.

[0025] In response to a change in network topology, routing engine 8 can dynamically reroute packets for multiple routes by changing a common portion of next hop data 18. More specifically, because routes using the same next hops share a common portion of next hop data 18, routing engine 8 can update next hop data 18 without needing to update route data 14, which can be significantly large for some networks. In this fashion, routing engine 8 can update large number of routes, and thereby quickly reroute packets, with minimal changes to the routing information 10.

[0026] FIG. 2A is a block diagram illustrating example data structures for one arrangement of routing information 10. In the illustrated embodiment, route data 14 of routing information 10 is arranged as a radix tree 19 that maps network routes to indirect next hop data 16 and next hop data 18. More specifically, radix tree 19 includes a number of leaf nodes 22A, 22B, 22C, 22D, collectively referred to as leaf nodes 22. Each of leaf nodes 22 corresponds to a network route. For large networks, radix tree 19 can become sizable and may easily include over 300,000 leaf nodes 22. Consequently, for exemplary purposes, FIG.2 depicts a portion of radix tree 19. The arrangement of routing information 10 as a radix tree is illustrated for exemplary purposes. The principles of invention may readily be applied to other arrangements. Routing information 10 may be arranged, for example, as a number of tables, link lists, and other data structures that store pointers to indirect next hop data 16 and next hop data 18.

[0027] Upon receiving an inbound packet, routing engine 8 reads a block of data corresponding to the packet, referred to as the “key,” that includes a network destination. The key may, for example, contain a routing prefix for another router within the network.

Routing engine 8 resolves the key to one of leaf nodes 22 by traversing radix tree 19. In particular, routing engine 8 traverses radix tree 19 by sequentially testing bits A, B and C, which represent any bits within the key. Based on the value of each bit A, B, C, routing engine 8 follows the links of radix tree 19 through the various levels until reaching one of leaf nodes 22.

[0028] Leaf nodes 22A, 22B and 22C include indirect next hop data 16 that references an array 24 storing next hop data 18. In particular, the indirect next hop data 16 points to one of the elements of array 24, thereby identifying a corresponding next hop for a respective network destination. In this manner, leaf nodes 22A, 22B and 22C of radix tree 19 do not contain next hop information, but include references to next hop data 18 that is stored in a separate data structure. In this fashion, indirect next hop data 16 provides intermediate data structures that relate route data 14 to next hop data 18. Leaf node 22D stores, however, stores next hop data NH12, and thereby bypasses indirect next hop data 16. In an alternative embodiment, array 24 may store references to specific interface ports, processing modules, or both.

[0029] Upon resolving a key of an inbound packet to one of leaf nodes 22A, 22B and 22C, routing engine 8 uses the contained one of indirect next hop data 16 to read next hop data from the referenced element of array 24. In the illustrated example, routing engine 8 resolves a packet key of "010" to leaf node 22C. Routing engine 8 uses the pointer contained within indirect hop data structure 22C to access the fourth element of array 24, i.e., the element with an index equal to 3, thereby resolving the key to next hop data NH2. Upon resolving the destination to a next hop, routing engine 8 determines an interface port associated with the actual next hop NH2, and forwards the inbound packet to the appropriate IFC 6 for transmission.

[0030] As illustrated in FIG. 2A, network routes corresponding to leaf nodes 22A, 22B, 22C share a common next hop. In other words, router 4 forwards all packets destined for these routes to the same neighboring network node, i.e., the same next hop. Consequently, according to the principles of the invention, the indirect next hop data 16 within leaf nodes 22A, 22B, 22C reference a common portion of next hop data, i.e., element 3 of array 24. If a network event occurs that requires rerouting packets along these routes, such as failure of the link between router 4 and the neighboring device, routing engine 8 can dynamically reroute

the packets by modifying array 24. In particular, routing engine 8 can overwrite the next hop data NH2 of element 3 with new next hop data. In response to a network event, routing engine 8 may, for example, write NH12 to element 3 of array 24, thereby quickly rerouting packets destined for NH2 to an alternate next hop, i.e., NH12.

[0031] In this manner, separating route data 14 from next hop data 18 by indirect next hop data 16 provides many advantages. Routing engine 8, for example, need not update radix tree 19 and, in particular, each of leaf nodes 22A, 22B and 22C. In large networks, it is not uncommon for 50,000 or more network destinations to have the same next hop from a routing device. By making use of intermediate references between radix tree 19 and the next hop data stored within array 18, instead of incorporating the next hop data within the radix tree 19, routing engine 8 need not change all of the affected leaf nodes 22, only the common next hop data. In this fashion, router 4 can dynamically reroute packets with minimal changes to the routing information 10.

[0032] FIG. 2B is a block diagram illustrating a second exemplary arrangement of routing information 10. In the illustrated embodiment, route data 14 of routing information 10 is arranged as a radix tree 40 that maps network routes to indirect next hop data 16 and next hop data 18. Radix tree 40 includes a number of leaf nodes 32A, 32B, 32C, 32D, collectively referred to as leaf nodes 32.

[0033] As illustrated in FIG. 2B, indirect next hop data 16 of leaf nodes 32 may include multiple references to array 34 storing next hop data 18. Leaf nodes 32A and 32B, for example, include primary (P) references and backup (B) references to portions of next hop data 18. For exemplary purposes, the primary references of leaf nodes 32A and 32B reference a common portion of next hop data, i.e., element 3 of array 24. The backup references of leaf nodes 32A and 32B reference a common backup next hop, as identified within element 0 of array 24. In this manner, indirect next hop data 16 for nodes 32A, 32B indicates that next hop NH5 is to be used in the event next hop NH2 fails. Leaf node 32C includes a single primary reference that identifies a portion of next hop data 18.

[0034] To generate radix tree 40, routing engine 8 precomputes alternative next hop and adds the alternative next hops to array 34. Based on the alternative next hops, routing engine 8 may include primary and backup references within leaf nodes 32 of radix tree 40, and may mark these references as active or inactive based on the current network topology.

[0035] This arrangement may provide a number of advantages when a network event, such as failure of a network link, requires router 4 to reroute packets. In particular, routing engine 8 can quickly reroute the packets from a primary next hop to a backup next hop without regenerating radix tree 40. For routes making use of a failed next hop, routing engine 8 may promote any backup references within corresponding leaf nodes 32 to primary references, and may mark the existing primary references as inactive. In this manner, routing engine 8 can quickly reroute packets to a precomputed backup next hop with minimal changes to routing information 10.

[0036] When an inactive next hop becomes available, routing engine 8 identifies those leaf nodes 32 referencing the next hop and marks the references as active. Routing engine 8 may promote the newly activated next hop to a primary next hop, or may designate the next hop as a backup next hop.

[0037] FIG. 3 is a block diagram illustrating an example data structure 70 for resolving next hop data 18 to interface ports. In the illustrated embodiment, data structure 70 forms a two-dimensional array having N rows and 3 columns. Each row uniquely associates a next hop with an interface port. Row 0, for example, associates next hop data NH10 with interface port IFC10. Although as illustrated each row maps a next hop to an interface port, data structure 70 could be used to map a next hop to any type of processing module. A row may, for example, map a packet to one of a number of network protocol modules, such as TCP/IP or MPLS, executing on the router for processing. In addition, a row may list zero or more other processing modules including, for example, a packet filtering module, a packet counting module, a policy enforcement module, and a rate-limiting module.

[0038] FIG. 4 is a block diagram illustrating another example router 74 configured consistent with the principles of the invention. Router 74 includes control unit 78 that directs inbound packets received from inbound link 12 to the appropriate outbound link 13. In particular, the functionality of control unit 78 is divided between a routing engine 85 and a packet forwarding engine 80.

[0039] Routing engine 85 is primarily responsible for maintaining routing information 86 to reflect the current network topology. In particular, routing engine 85 periodically updates routing information 86 to accurately reflect the network topology.

[0040] In accordance with routing information 86, packet forwarding engine 85 maintains forwarding information 84 that associates network destinations with specific next hops and corresponding interface ports of IFCs 6. Forwarding information 84 may, therefore, be thought of as a subset of the information contained within routing information 86. Upon receiving an inbound packet, forwarding engine 84 directs the inbound packet to an appropriate IFCs 6 for transmission based on forwarding information 84. In one embodiment, each of packet forwarding engine 80 and routing engine 85 may comprise one or more dedicated processors, hardware, and the like, and may be communicatively coupled by data communication channel 82. Data communication channel 82 may be a high-speed network connection, bus, shared-memory or other data communication mechanism.

[0041] When a network event occurs, such as a link failure, routing engine 85 updates routing information 86 and directs packet forwarding engine 80 to update forwarding information 84. Routing engine 85 may, for example, communicate one or more messages over data communication channel 82 directing packet forwarding engine 80 to update the next hop data for one or more network destinations.

[0042] Forwarding engine 80, routing engine 85, or both, may make use of the data structures and organization described above. In particular, packet forwarding engine 80 may maintain forwarding information 84 so as to make use of indirect next hop data. The indirect next hop data may associate, for example, leaf nodes of a forwarding tree with next hop data. This embodiment may be advantageous in that, in response to an update message from routing engine 85, packet forwarding engine 80 may need only update next hop data that is referenced by the indirect next hop data structures, and not the forwarding tree itself.

[0043] In addition, routing engine 85 may organize routing information 86 to include a local copy 90 of forwarding information 84, or portions thereof. This embodiment may be particularly advantageous in reducing the number of messages between routing engine 85 and packet forwarding engine 80. Upon updating routing information 86 due to a change of network topology, routing engine 85 may identify the next hop data to be changed by examining the copy of forwarding information 90. Based on the examination, routing engine 85 may generate a limited number of messages directing forwarding engine 85 to appropriately update next hop data within forwarding information 84. In particular, routing engine 85 may generate a single message directing packet forwarding engine 80 to overwrite

a common next hop datum referenced by indirect next hop data structures within leaf nodes of a forwarding tree. This may greatly reduce the number of messages between routing engine 85 and packet forwarding engine 80, primarily because the number of messages is no longer a function of the number of routes affected by the change, as with conventional routers.

[0044] FIG. 5 is a flow chart illustrating an example operation of routing engine 85 consistent with the principles of the invention. Upon receiving new network information from another network node (94), routing engine 85 updates routing information 86. Router 4 may receive, for example, network information via the Border Gateway Protocol (BGP) or other protocol for sharing network information.

[0045] After updating routing information 86, routing engine 85 examines a local copy of forwarding information 90 to determine whether packets can be dynamically rerouted by changing next hop data (98). Based on the examination, routing engine 85 may generate one or more update messages directing packet forwarding engine 80 to appropriately update next hop data within forwarding information 84 (100). Routing engine 85 communicates the messages to packet forwarding engine 80 to update next hop data as referenced by indirect next hop data structures within leaf nodes of a forwarding tree within forwarding information 84. As described above, packet forwarding engine 80 may dynamically reroute packets by modifying a common portion of the next hop data, such as an element of an array.

Forwarding engine 80 may, for example, write new next hop to one or more elements of the array, thereby quickly rerouting packets to an alternate next hop. In this manner, packet forwarding engine 80 may not need to change the forwarding tree itself, which may greatly reduce the number of messages between routing engine 85 and forwarding engine 80.

[0046] Furthermore, packet forwarding engine 80 may promote backup next hop references within corresponding leaf nodes of the forwarding tree to primary next hop references, and may mark existing primary next hop references as inactive. In this manner, packet forwarding engine 80 may quickly reroute packets to a precomputed backup next hop.

[0047] Various embodiments of the invention have been described that provide for increased efficiency in updating routing information after a change in network topology, such as link failure. These and other embodiments are within the scope of the following claims.